UDP-N-Acetylglucosamine: Galactose-β1,3-N-Acetylgalactosamine-α-R / N-Acetylglucosamine-β1,3-N-Acetylgalactosamine-α-R (GlcNAc to GalNAc) β1,6-N-Acetylglucosaminyltransferase, C2/4GnT

INS A1

TECHNICAL FIELD

The present invention relates generally to the biosynthesis of glycans found as free oligosaccharides or covalently bound to proteins and glycolipids. This invention is more particularly related to a family of nucleic acids encoding UDP-N-acetylglucosamine: N-acetylgalactosamine β1,6-N-acetylglucosaminyltransferases (Core-β1,6-N-acetylglucosaminyltransferases), which add N-acetylglucosamine to the hydroxy group at C6 of 2-acetamido-2-deoxy-D-galactosamine (GalNAc) in O-glycans of the core 3 and the core 1 type. This invention is more particularly related to a gene encoding the third member of the family of O-glycan β1,6-N-acetylglucosaminyltransferases, termed C2/4GnT, probes to the DNA encoding C2/4GnT, DNA constructs comprising DNA encoding C2/4GnT, recombinant plasmids and recombinant methods for producing C2/4GnT, recombinant methods for stably transforming or transfecting cells for expression of C2/4GnT, and methods for identification of DNA polymorphism in patients.

BACKGROUND OF THE INVENTION

O-linked protein glycosylation involves an initiation stage in which a family of N-acetylgalactosaminyltransferases catalyzes the addition of *N*-acetylgalactosamine to serine or threonine residues (1). Further assembly of *O*-glycan chains involves several sucessive or alternative biosynthetic reactions: i) formation of simple mucin-type core 1 structures by UDP-Gal: GalNAcα-R β1,3Gal-transferase activity; ii) conversion of core 1 to complex-type core 2 structures by UDP-GlcNAc: Galβ1-3GalNAcα-R β1,6GlcNAc-transferase activities; iii) direct formation of complex mucin-type core 3 by UDP-GlcNAc: GalNAcα β1,3GlcNAc-transferase activities; and iv) conversion of core 3 to core 4 by UDP-GlcNAc: GlcNAcβ1-3GalNAcα-R β1,6GlcNAc-transferase activity. The formation of 1,6GlcNAc branches (reactions ii and iv) may be considered a key controlling event of *O*-linked protein glycosylation leading to structures produced upon differentiation and

malignant transformation (2-6). For example, increased formation of GlcNAcβ1-6GalNAc branching in O-glycans has been demonstrated during T-cell activation, during the development of leukemia, and for immunodeficiencies like Wiskott-Aldrich syndrome and AIDS (7; 8). Core 2 branching may play a role in tumor progression and metastasis (9). In contrast, many carcinomas show changes from complex O-glycans found in normal cell types to immaturely processed simple mucin-type O-glycans such as T (Thomsen-Friedenreich antigen; Gal 1-3GalNAc 1-R), Tn (GalNAc 1-R), and sialosyl-Tn (NeuAc 2-6GalNAc 1-R) (10). The molecular basis for this has been extensively studied in breast cancer, where it was shown that specific downregulation of core 2 β6GlcNAc-transferase was responsible for the observed lack of complex type *O*-glycans on the mucin MUC1 (6). *O*-glycan core assembly may therefore be controlled by inverse changes in the expression level of Core-β1,6-N-acetylglucosaminyltransferases and the sialyltransferases forming sialyl-T and sialyl-Tn .

Interestingly, the metastatic potential of tumors has been correlated with increased expression of core 2 β6GlcNAc-transferase activity (5). The increase in core 2 β6GlcNAc-transferase activity was associated with increased levels of poly *N*-acetyllactosamine chains carrying sialyl-Le^x, which may contribute to tumor metastasis by altering selectin mediated adhesion (4; 11). The control of *O*-glycan core assembly is regulated by the expression of key enzyme activities outlined in Figure 1; however, epigenetic factors including posttranslational modification, topology, or competition for substrates may also play a role in this process (11).

The *in vitro* biosynthesis of a subset of complex *O*-glycopeptide structures is presently hampered by lack of availability of the enzymes adding *N*-acetylglucosamine in a β1-3 linkage to GalNAcα1-*O*-Ser/Thr to form core 3 as well as the enzyme catalyzing the successive addition of β1-6 *N*-acetylglucosamine branches to form core 4. This structure is required for the enzymes responsible for further build-up of core 4 based complex type *O*-glycans (Fig. 1). Most other enzymes required for elongation of branched *O*-glycans are available, and the core 2/4 enzyme described herein now makes the synthesis of core 4 based structures possible.

Access to the gene encoding C2/4GnT would allow production of a glycosyltransferase for use in formation of core 2 or core 4 – based O-glycan modifications on oligosacccharides, glycoproteins and glycosphingolipids. This enzyme could be used, for example in pharmaceutical or other commercial

applications that require synthetic addition of core 2 or core 4 based O-glycans to these or other substrates, in order to produce appropriately glycosylated glycoconjugates having particular enzymatic, immunogenic, or other biological and/or physical properties.

Consequently, there exists a need in the art for UDP-N-Acetylglucosamine: Galactose-β1,3-N-Acetylgalactosamine-α-R / N-Acetylglucosamine-β1,3-N-Acetylgalactosamine-α-R (GlcNAc to GalNAc) β1-6 N-Acetylglucosaminyltransferase and the primary structure of the gene encoding these enzyme. The present invention meets this need, and further presents other related advantages.

SUMMARY OF THE INVENTION

The present invention provides isolated nucleic acids encoding human UDP-N-acetylglucosamine: N-acetylgalactosamine β1,6 N-acetylglucosaminyltransferasee (C2/4GnT), including cDNA and genomic DNA. C2/4GnT has broader acceptor substrate specificities compared to C2GnT, as exemplified by its activity with core 3- -R saccharide derivatives. The complete nucleotide sequence of C2/4GnT is set forth in SEQ ID NO:1 and Figure 2.

In one aspect, the invention encompasses isolated nucleic acids comprising the nucleotide sequence of nucleotides 496-1812 as set forth in SEQ ID NO:1 and Figure 2 or sequence-conservative or function-conservative variants thereof. Also provided are isolated nucleic acids hybridizable with nucleic acids having the sequence as set forth in SEQ ID NO:1 and Figure 2 or fragments thereof or sequence-conservative or function-conservative variants thereof; preferably, the nucleic acids are hybridizable with C2/4GnT sequences under conditions of intermediate stringency, and, most preferably, under conditions of high stringency. In one embodiment, the DNA sequence encodes the amino acid sequence shown in SEQ ID NO:2 and Figure 2 from methionine (amino acid no. 1) to leucine (amino acid no. 438). In another embodiment, the DNA sequence encodes an amino acid sequence comprising a sequence from phenylalanine (no. 31) to leucine (no.438) of the amino acid sequence set forth in SEQ ID NO:2 and Figure 2.

In a related aspect, the invention provides nucleic acid vectors comprising C2/4GnT DNA sequences, including but not limited to those vectors in which the C2/4GnT DNA sequence is operably linked to a transcriptional regulatory element, with or without a polyadenylation sequence. Cells comprising these vectors are also provided, including

without limitation transiently and stably expressing cells. Viruses, including bacteriophages, comprising C2/4GnT-derived DNA sequences are also provided. The invention also encompasses methods for producing C2/4GnT polypeptides. Cell-based methods include without limitation those comprising: introducing into a host cell an isolated DNA molecule encoding C2/4GnT, or a DNA construct comprising a DNA sequence encoding C2/4GnT; growing the host cell under conditions suitable for C2/4GnT expression; and isolating C2/4GnT produced by the host cell. A method for generating a host cell with *de novo* stable expression of C2/4GnT comprises: introducing into a host cell an isolated DNA molecule encoding C2/4GnT or an enzymatically active fragment thereof (such as, for example, a polypeptide comprising amino acids 31-438 of the amino acid sequence set forth in SEQ ID NO:2 and Figure 2), or a DNA construct comprising a DNA sequence encoding C2/4GnT or an enzymatically active fragment thereof; selecting and growing host cells in an appropriate medium; and identifying stably transfected cells expressing C2/4GnT. The stably transfected cells may be used for the production of C2/4GnT enzyme for use as a catalyst and for recombinant production of peptides or proteins with appropriate galactosylation. For example, eukaryotic cells, whether normal or diseased cells, having their glycosylation pattern modified by stable transfection as above, or components of such cells, may be used to deliver specific glycoforms of glycopeptides and glycoproteins, such as, for example, as immunogens for vaccination.

In yet another aspect, the invention provides isolated C2/4GnT polypeptides, including without limitation polypeptides having the sequence set forth in SEQ ID NO:2 and Figure 2, polypeptides having the sequence of amino acids 31-438 as set forth in SEQ ID NO:2 and Figure 2, and a fusion polypeptide consisting of at least amino acids 31-438 as set forth in SEQ ID NO:2 and Figure 2 fused in frame to a second sequence, which may be any sequence that is compatible with retention of C2/4GnT enzymatic activity in the fusion polypeptide. Suitable second sequences include without limitation those comprising an affinity ligand or a reactive group.

In another aspect of the present invention, methods are disclosed for screening for mutations in the coding region (exon III) of the C2/4GnT gene using genomic DNA isolated from, e.g., blood cells of patients. In one embodiment, the method comprises: isolation of DNA from a patient; PCR amplification of coding exon III; DNA sequencing of amplified exon DNA fragments and establishing therefrom potential structural defects of the C2/4GnTgene associated with disease.

These and other aspects of the present invention will become evident upon reference to the following detailed description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the biosynthetic pathways of mucin-type O-glycan core structures. The abbreviations used are GalNAc-T: polypeptide αGalNAc-transferase; ST6GalNAcI: mucin α2,6 sialyltransferase; C1β3Gal-T: core 1 β1,3 galactosyltransferase; C2GnT: core 2 β1,6 GlcNAc-transferase; C2/4GnT: core2 / core 4 β1,6 GlcNAc-transferase; C3GnT: core 3 β1,3 GlcNAc-transferase; ST3GalI: mucin α2,3 sialyltransferase; β4Gal-T: β1,4 galactosyltransferase; β3Gal-T: β1,3 galactosyltransferase; β3GnT: elongation β1,3 GlcNAc-transferase.

Sub B1

Figure 2 depicts the DNA sequence of the C2/4GnT (accession # AF038650) gene and the predicted amino acid sequence of C2/4GnT. The amino acid sequence is shown in single letter code. The hydrophobic segment representing the putative transmembrane domain is double underlined. Two consensus motifs for N-glycosylation are indicated by *asterisks*. The location of the primers used for preparation of the expression constructs are indicated by *single underlining*. A potential polyadenylation signal is indicated in *boldface underlined type*.

Sub B2

Figure 3 is an illustration of a sequence comparison between human C2GnT (accession # M97347), human C2/4GnT (accession # AF038650), and human I-GnT (accession # Z19550). Introduced gaps are shown as *hyphens*, and aligned identical residues are *boxed* (*black* for all sequences, and *grey* for two sequences). The putative transmembrane domains are *underlined* with a *single line*. The positions of conserved cysteines are indicated by *asterisks*. One conserved *N*-glycosylation sites is indicated by an *open circle*.

Figure 4 depicts a Northern blot analysis of healthy human tissues and gastric cancer cell lines. **Panel A:** Multiple human tissue northern blots, MTN I and MTN II, from Clontech were probed with a ^{32}P-labeled probe corresponding to the soluble expression fragment of C2/4GnT (base pairs 91-1317). **Panel B:** A northern blot of total RNA from human colonic and pancreatic cancer cell lines was probed as described for panel A.

Figure 5 depicts sections of a 1-D 1H-NMR spectrum of the C2/4GnT product. GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα1-1-*p*Nph, showing all non-exchangeable monosaccharide ring methine and exocyclic methylene resonances. Residue designations for GlcNAcβ1→3 (β3), GlcNAcβ1→6 (β6), and GalNAcα1→1 (α) are followed by proton designations (1-6). All resonances in this region except for β3-5 (3.453 ppm) are marked.

Figure 6 is a section of the ^{1}H-detected ^{1}H-^{13}C heteronuclear multiple bond correlation (HMBC) spectrum of the Core 4 β6 GlcNAc transferase product, showing interglycosidic H1–C1–O1–Cx and C1–O1–Cx–Hx correlations (cross-peaks marked by ovals). The unmarked cross-peaks are all intra-residue correlations.

Figure 7 shows a fluorescence *in situ* hybridization of C2/4GnT to metaphase chromosomes. The C2/4GnT probe (P1 DNA from clone DPMC-HFF#1-1091[F1]) labeled band 15q21.3

Figure 8 is a schematic representation of forward (TSHC78) and reverse (TSHC79) PCR primers that can be used to amplify the coding exon of the C2/4GnT gene. The sequences of the primers are also shown. TSHC78 has SEQ ID NO:9 and TSHC79 has SEQ ID NO:10.

DETAILED DESCRIPTION OF THE INVENTION

All patent applications, patents, and literature references cited in this specification are hereby incorporated by reference in their entirety. In the case of conflict, the present description, including definitions, is intended to control.

Definitions:

1. "Nucleic acid" or "polynucleotide" as used herein refers to purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribo nucleotides. This includes single- and double-stranded molecules, i.e., DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases (see below).

2. "Complementary DNA or cDNA" as used herein refers to a DNA molecule or sequence that has been enzymatically synthesized from the sequences present in a mRNA

template, or a clone of such a DNA molecule. A "DNA Construct" is a DNA molecule or a clone of such a molecule, either single- or double-stranded, which has been modified to contain segments of DNA that are combined and juxtaposed in a manner that would not otherwise exist in nature. By way of non-limiting example, a cDNA or DNA which has no introns is inserted adjacent to, or within, exogenous DNA sequences.

3. A plasmid or, more generally, a vector, is a DNA construct containing genetic information that may provide for its replication when inserted into a host cell. A plasmid generally contains at least one gene sequence to be expressed in the host cell, as well as sequences that facilitate such gene expression, including promoters and transcription initiation sites. It may be a linear or closed circular molecule.

4. Nucleic acids are "hybridizable" to each other when at least one strand of one nucleic acid can anneal to another nucleic acid under defined stringency conditions. Stringency of hybridization is determined, e.g., by a) the temperature at which hybridization and/or washing is performed, and b) the ionic strength and polarity (e.g., formamide) of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two nucleic acids contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stringency (such as, for example, in an aqueous solution of 0.5X SSC, at 65 °C) requires that the sequences exhibit some high degree of complementarity over their entire sequence. Conditions of intermediate stringency (such as, for example, an aqueous solution of 2X SSC at 65 °C) and low stringency (such as, for example, an aqueous solution of 2X SSC at 55 °C), require correspondingly less overall complementarily between the hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate.)

5. An "isolated" nucleic acid or polypeptide as used herein refers to a component that is removed from its original environment (for example, its natural environment if it is naturally occurring). An isolated nucleic acid or polypeptide contains less than about 50%, preferably less than about 75%, and most preferably less than about 90%, of the cellular components with which it was originally associated.

6. A "probe" refers to a nucleic acid that forms a hybrid structure with a sequence in a target region due to complementarily of at least one sequence in the probe with a sequence in the target region.

7. A nucleic acid that is "derived from" a designated sequence refers to a nucleic acid sequence that corresponds to a region of the designated sequence. This encompasses sequences that are homologous or complementary to the sequence, as well as "sequence-conservative variants" and "function-conservative variants". Sequence-conservative variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position. Function-conservative variants of C2/4GnT are those in which a given amino acid residue in the polypeptide has been changed without altering the overall conformation and enzymatic activity (including substrate specificity) of the native polypeptide; these changes include, but are not limited to, replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and the like).

8. A "donor substrate" is a molecule recognized by, e.g., a Core-β1,6-N-acetyl-glucosaminyltransferase and that contributes an N-acetylglucosaminyl moiety for the transferase reaction. For C2/4GnT, a donor substrate is UDP-N-acetylglucosamine. An "acceptor substrate" is a molecule, preferably a saccharide or oligosaccharide, that is recognized by, e.g., an N-acetylglucosaminyltransferase and that is the target for the modification catalyzed by the transferase, i.e., receives the N-acetylglucosaminyl moiety. For C2/4GnT, acceptor substrates include without limitation oligosaccharides, glycoproteins, O-linked core 1- and core 3-glycopeptides, and glycosphingolipids comprising the sequences Gal 1-3GalNAc, GlcNAc 1-3GalNAc or Glc 1-3GalNAc.

The present invention provides the isolated DNA molecules, including genomic DNA and cDNA, encoding the UDP-N-acetylglucosamine: N-acetylgalactosamine 1,6 N-acetylglucosaminyltransferase (C2/4GnT).

C2/4GnT was identified by analysis of EST database sequence information, and cloned based on EST and 5'RACE cDNA clones. The cloning strategy may be briefly summarized as follows: 1) synthesis of oligonucleotides derived from EST sequence information, designated TSHC27 (SEQ ID NO:3) and TSHC28 (SEQ ID No.4); 2) successive 5'-rapid amplification of cDNA ends (5'RACE) using commercial Marathon-Ready cDNA; 3) cloning and sequencing of 5'RACE cDNA; 4) identification of a novel cDNA sequence corresponding to C2/4GnT; 5) construction of expression constructs by reverse-transcription-polymerase chain reaction (RT-PCR) using Colo205 human cell line mRNA; 6) expression of the cDNA encoding C2/4GnT in Sf9 (*Spodoptera frugiperda*) cells. More specifically, the isolation of a representative DNA molecule encoding a novel

second member of the mammalian UDP-N-acetylglucosamine: β-*N*-actylgalactosamine β1,6-N-acetylglucosaminyltransferase family involved the following procedures described below.

Identification of DNA homologous to C2GnT.

Database searches were performed with the coding sequence of the human C2GnT sequence (12) using the BLASTn and tBLASTn algorithms against the dbEST database at The National Center for Biotechnology Information, USA. The BLASTn algorithm was used to identify ESTs representing the query gene (identities of 95%), whereas tBLASTn was used to identify non-identical, but similar EST sequences. ESTs with 50-90% nucleotide sequence identity were regarded as different from the query sequence. One EST with several apparent short sequence motifs and cysteine residues arranged with similar spacing was selected for further sequence analysis.

Cloning of human C2/4GnT.

EST clone 178656 (5' EST GenBank accession number AA307800), derived from a putative homologue to C2GnT, was obtained from the American Type Culture Collection, USA. Sequencing of this clone revealed a partial open reading frame with significant sequence similarity to C2GnT. The coding region of human C2GnT and a bovine homologue was previously found to be organized in one exon ((13), and unpublished observations). Since the 5' and 3' sequence available from the C2/4GnT EST was incomplete but likely to be located in a single exon, the missing 5' and 3' portions of the open reading frame was obtained by sequencing genomic P1 clones. P1 clones were obtained from a human foreskin genomic P1 library (DuPont Merck Pharmaceutical Co. Human Foreskin Fibroblast P1 Library) by screening with the primer pair TSHC27 (5'-GGAAGTTCATACAGTTCCCAC-3') (SEQ ID NO:3) and TSHC28 (5'-CCTCCCATTCAACATCTTGAG -3') (SEQ ID NO:4). Two genomic clones for C2/4GnT, DPMC-HFF#1-1026(E2) and DPMC-HFF#1-1091(F1) were obtained from Genome Systems Inc. DNA from P1 phage was prepared as recommended by Genome Systems Inc. The entire coding sequence of the C2/4GnT gene was represented in both clones and sequenced in full using automated sequencing (ABI377, Perkin-Elmer). Confirmatory sequencing was performed on a cDNA clone obtained by PCR (30 cycles at 95 °C for 15 sec; 55 °C for 20 sec and 68 °C for 2 min 30 sec) on total cDNA from the human COLO 205 cancer cell line with the sense primer TSHC54 (5'- GCAGAATTCATGGTTCAATGGAAGAGACTC-3')

(SEQ ID NO:7) and the anti-sense primer TSHC45 (5'- AGCGAATTCAGCTCAAAGTTCAGTCCCATAG -3') (SEQ ID NO:5). The composite sequence contained an open reading frame of 1314 base pairs encoding a putative protein of 438 amino acids with type II domain structure predicted by the TMpred-algorithm at the Swiss Institute for Experimental Cancer Research (ISREC) (http://www.isrec.isb-sib.ch/software/TMPRED_form.html). The sequence of the 5'-end of C2/4GnT mRNA including the translational start site and 5'-UTR was obtained by 5' rapid amplification of cDNA ends (35 cycles at 94 °C for 20 sec; 52 °C for 15 sec and 72 °C for 2 min) using total cDNA from the human COLO 205 cancer cell line with the anti-sense primer TSHC48 (5'- GTGGGAACTGTATGAACTTCC-3') (SEQ ID NO:6) (Fig. 2).

Expression of C2/4GnT.

An expression construct designed to encode amino acid residues 31-438 of C2/4GnT was prepared by PCR using P1 DNA, and the primer pair TSHC55 (5'-CGAGAATTCAGGTTGAAGTGTGACTC -3') (SEQ ID NO:8) and TSHC45 (SEQ ID NO:5) (Fig. 2). The PCR product was cloned into the *Eco*RI site of pAcGP67A (PharMingen), and the insert was fully sequenced. pAcGP67-C2/4GnT-sol was co-transfected with Baculo-Gold™ DNA (PharMingen) as described previously (14). Recombinant Baculo-virus were obtained after two successive amplifications in Sf9 cells grown in serum-containing medium, and titers of virus were estimated by titration in 24-well plates with monitoring of enzyme activities. Transfection of Sf9-cells with pAcGP67-C2/4GnT-sol resulted in marked increase in GlcNAc-transferase activity compared to uninfected cells or cells infected with a control construct. C2/4GnT showed significant activity with disaccharide derivatives of *O*-linked core 1 (Galβ1-3GalNAcα1-R) and core 3 structures (GlcNAcβ1-3GalNAcα1-R). In contrast, no activity was found with lacto-*N*-*neo*tetraose as well as GlcNAcβ1-3Gal-Me as acceptor substrates indicating that C2/4GnT has no IGnT-activity. Additionally, no activity could be detected wih α-D-GalNAc-1- *para*-nitrophenyl indicating that C2/4GnT does not form core 6 (GlcNAcβ1-6GalNAcα1-R) (Table I). No substrate inhibition of enzyme activity was found at high acceptor concentrations up to 20 mM core1- *para*-nitrophenyl or core3- *para*-nitrophenyl. C2/4GnT shows strict donor substrate specificity for UDP-GlcNAc, no activity could be detected with UDP-Gal or UDP-GalNAc (data not shown).

Table I: Substrate specificities of C2/4GnT and C2GnT

Substrate	C2/4GnT[a]		C2GnT	
	2 mM	10 mM	2 mM	10 mM
	nmol / h / mg		*nmol / h / mg*	
β-D-Gal-(1-3)-α-D-GalNAc	2.8	7.3	9.6	19.0
β-D-Gal-(1-3)-α-D-GalNAc-1-*p*-Nph	16.1	21.8	16.2	23.6
β-D-GlcNAc-(1-3)-α-D-GalNAc-1-*p*-Nph	5.2	7.4	<0.1	<0.1
α-D-GalNAc-1-*p*-Nph	<0.1	<0.1	<0.1	<0.1
D-GalNAc	<0.1	<0.1	<0.1	<0.1
lacto-*N*-*neo*-tetraose	<0.1	<0.1	<0.1	<0.1
β-D-GlcNAc-(1-3)-β-D-Gal-1-Me	<0.1	<0.1	<0.1	<0.1

[a] Enzyme sources were partially purified media of infected High Five™ cells (see "Experimental Procedures"). Background values obtained with uninfected cells or cells infected with an irrelevant construct were subtracted. [b] Me, methyl; Nph, nitrophenyl.

Controls included the pAcGP67-GalNAc-T3-sol (15). The kinetic properties were determined with partially purified enzymes expressed in High Five™ cells. Partial purification was performed by consecutive chromatography on Amberlite IRA-95, DEAE-Sephacryl and CM-Sepharose essentially as described (16).

__Northern blot analysis of human organs.__

Human multiple tissue northern blots containing mRNA from healthy human adult organs (Clontech) were probed with a C2/4GnT-probe. Northern analysis with mRNA from sixteen organs showed expression of C2/4GnT in organs of the gastrointestinal tract with high transcription levels observed in colon and kidney and lower levels in small intestine and pancreas (Fig. 4A). To investigate changes in expression of C2/4GnT in cancer cells derived from tissues normally expressing C2/4GnT, mRNA levels in a panel of human adenocarcinoma cell lines were determined. Analyses of C2/4GnT transcription levels revealed differential expression in pancreatic cell lines: Capan-1 and AsPC-1 expressed the transcript, whereas PANC-1, Capan-2, and BxPC-3 did not (Fig. 4B). Of the colonic cell lines, only HT-29 expressed transcripts of C2/4GnT. The size of the predominant transcript was approximately 2.4 kilobases, which correlates to the transcript size of 2.1 kilobases of the smallest of three transcripts of human C2GnT (12). Additionally, transcripts of approximately 3.4 kilobases and 6 kilobases were obtained in mRNA from

healthy colonic mucosa (Fig. 4A). The two additional transcripts may resemble the 3.3 kilobase and 5.4 kilobase transcripts of C2GnT, which have not yet been characterized. Multiple transcripts of C2GnT have been suggested to be caused by differential usage of polyadenylation signals, which affects the length of the 3' UTR (12).

Genomic organization of C2/4GnT gene.

The present invention also provides isolated genomic DNA molecules encoding C2/4GnT. A human genomic foreskin P1 library (DuPont Merck Pharmaceutical Co. Human Foreskin Fibroblast P1 Library) by screening with the primer pair
TSHC27 (5'-GGAAGTTCATACAGTTCCCAC-3') (SEQ ID NO:3) and
TSHC28 (5'-CCTCCCATTCAACATCTTGAG -3') (SEQ ID NO:4),
located in the coding exon yielding a product of 400 bp. Two genomic clones for C2/4GnT, DPMC-HFF#1-1026(E2) and DPMC-HFF#1-1091(F1) were obtained from Genome Systems Inc. The P1 clone was partially sequenced and introns in the 5'-untranslated region of C2/4GnT mRNA identified as shown in Figure 6. All exon/intron boundaries identified conform to the GT-AG consensus rule.

Chromosomal localization of C2/4GnT gene.

The present invention also discloses the chromosomal localization of the C2/4GnT gene. Fluorescence *in situ* hybridization to metaphase chromosomes using the isolated P1 phage clone DPMC-HFF#1-1091(F1) showed a fluorescence signal at 15q21.3 (Figure 7; 20 metaphases evaluated). No specific hybridization was observed at any other chromosomal site.

The C2/4GnT gene is selectively expressed in organs of the gastrointestinal tract. The C2/4GnT enzyme of the present invention was shown to exhibit O-glycosylation capacity implying that the C2/4GnT gene is vital for correct/full O-glycosylation *in vivo* as well. A structural defect in the C2/4GnT gene leading to a deficient enzyme or completely defective enzyme would therefore expose a cell or an organism to protein/peptide sequences which were not covered by O-glycosylation as seen in cells or organisms with intact C2/4GnT gene. Described in Example 6 below is a method for scanning the coding exon for potential structural defects. Similar methods could be used for the characterization of defects in the non-coding region of the C2/4GnT gene including the promoter region.

DNA, Vectors, and Host Cells

In practicing the present invention, many conventional techniques in molecular biology, microbiology, recombinant DNA, and immunology, are used. Such techniques are well known and are explained fully in, for example, Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; *DNA Cloning: A Practical Approach*, Volumes I and II, 1985 (D.N. Glover ed.); *Oligonucleotide Synthesis*, 1984, (M.L. Gait ed.); *Nucleic Acid Hybridization*, 1985, (Hames and Higgins); *Transcription and Translation*, 1984 (Hames and Higgins eds.); *Animal Cell Culture*, 1986 (R.I. Freshney ed.); *Immobilized Cells and Enzymes* , 1986 (IRL Press); Perbal, 1984, *A Practical Guide to Molecular Cloning*; the series, *Methods in Enzymology* (Academic Press, Inc.); *Gene Transfer Vectors for Mammalian Cells*, 1987 (J. H. Miller and M. P. Calos eds., Cold Spring Harbor Laboratory); *Methods in Enzymology* Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively); *Immunochemical Methods in Cell and Molecular Biology*, 1987 (Mayer and Waler, eds; Academic Press, London); Scopes, 1987, *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.) and *Handbook of Experimental Immunology*, 1986, Volumes I-IV (Weir and Blackwell eds.).

The invention encompasses isolated nucleic acid fragments comprising all or part of the nucleic acid sequence disclosed herein as set forth in SEQ ID NO:1 and Figure 2. The fragments are at least about 8 nucleotides in length, preferably at least about 12 nucleotides in length, and most preferably at least about 15-20 nucleotides in length. The invention further encompasses isolated nucleic acids comprising sequences that are hybridizable under stringency conditions of 2X SSC, 55 C, to the nucleotide sequence set forth in SEQ ID NO:1 and Figure 2; preferably, the nucleic acids are hybridizable at 2X SSC, 65 °C; and most preferably, are hybridizable at 0.5X SSC, 65 °C.

The nucleic acids may be isolated directly from cells. Alternatively, the polymerase chain reaction (PCR) method can be used to produce the nucleic acids of the invention, using either chemically synthesized strands or genomic material as templates. Primers used for PCR can be synthesized using the sequence information provided herein and can further be designed to introduce appropriate new restriction sites, if desirable, to facilitate incorporation into a given vector for recombinant expression.

The nucleic acids of the present invention may be flanked by natural human regulatory sequences, or may be associated with heterologous sequences, including promoters,

enhancers, response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3'- noncoding regions, and the like. The nucleic acids may also be modified by many means known in the art. Non-limiting examples of such modifications include methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoroamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.). Nucleic acids may contain one or more additional covalently linked moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), intercalators (e.g., acridine, psoralen, etc.), chelators (e.g., metals, radioactive metals, iron, oxidative metals, etc.), and alkylators. The nucleic acid may be derivatized by formation of a methyl or ethyl phosphotriester or an alkyl phosphoramidate linkage. Furthermore, the nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent molecules, biotin, and the like.

According to the present invention, useful probes comprise a probe sequence at least eight nucleotides in length that consists of all or part of the sequence from among the sequences as set forth in Figure 2 or sequence-conservative or function-conservative variants thereof, or a complement thereof, and that has been labelled as described above.

The invention also provides nucleic acid vectors comprising the disclosed sequence or derivatives or fragments thereof. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts, and may be used for gene therapy as well as for simple cloning or protein expression.

Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g. antibiotic resistance, and one or more expression cassettes. The inserted coding sequences may be synthesized by standard methods, isolated from natural sources, or prepared as hybrids, etc. Ligation of the coding sequences to transcriptional regulatory elements and/or to other amino acid coding sequences may be achieved by known methods. Suitable host cells may be transformed/transfected/infected as appropriate by any suitable method including electroporation, $CaCl_2$ mediated DNA uptake, fungal infection, microinjection, microprojectile, or other established methods.

Appropriate host cells included bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Pichia pastoris, Hansenula polymorpha, Neurospora, SF9* cells, C129 cells, 293 cells, and CHO cells, COS cells, HeLa cells, and immortalized mammalian myeloid and lymphoid cell lines. Preferred replication systems include M13, ColE1, 2 , ARS, SV40, baculovirus, lambda, adenovirus, and the like. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced C2/4GnT derived peptides and polypeptides.

Advantageously, vectors may also include a transcription regulatory element (i.e., a promoter) operably linked to the C2/4GnT coding portion. The promoter may optionally contain operator portions and/or ribosome binding sites. Non-limiting examples of bacterial promoters compatible with *E. coli* include: β-lactamase (penicillinase) promoter; lactose promoter; tryptophan (trp) promoter; arabinose BAD operon promoter; lambda-derived P_l promoter and N gene ribosome binding site; and the hybrid tac promoter derived from sequences of the trp and lac UV5 promoters. Non-limiting examples of yeast promoters include 3-phosphoglycerate kinase promoter, glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter, galactokinase (GAL1) promoter, galactoepimerase (GAL10) promoter, (CUP) copper cch and alcohol dehydrogenase (ADH) promoter. Suitable promoters for mammalian cells include without limitation viral promoters such as that from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences and enhancer sequences which increase expression may also be included; sequences which cause amplification of the gene may also be desirable. Furthermore, sequences that facilitate secretion of the recombinant product from cells, including, but not limited to, bacteria, yeast, and animal cells, such as secretory signal sequences and/or prohormone pro region sequences, may also be included. These sequences are known in the art.

Nucleic acids encoding wild type or variant polypeptides may also be introduced into cells by recombination events. For example, such a sequence can be introduced into a cell, and thereby effect homologous recombination at the site of an endogenous gene or a sequence

with substantial identity to the gene. Other recombination-based methods such as nonhomologous recombinations or deletion of endogenous genes by homologous recombination may also be used.

The nucleic acids of the present invention find use, for example, as probes for the detection of C2/4GnT in other species or related organisms and as templates for the recombinant production of peptides or polypeptides. These and other embodiments of the present invention are described in more detail below.

Polypeptides and Antibodies

The present invention encompasses isolated peptides and polypeptides encoded by the disclosed genomic sequence. Peptides are preferably at least five residues in length.

Nucleic acids comprising protein-coding sequences can be used to direct the recombinant expression of polypeptides in intact cells or in cell-free translation systems. The known genetic code, tailored if desired for more efficient expression in a given host organism, can be used to synthesize oligonucleotides encoding the desired amino acid sequences. The phosphoramidite solid support method of Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* **103**:3185, the method of Yoo *et al.*, 1989, *J. Biol. Chem.* **764**:17078, or other well known methods can be used for such synthesis. The resulting oligonucleotides can be inserted into an appropriate vector and expressed in a compatible host organism.

The polypeptides of the present invention, including function-conservative variants of the sequence disclosed in SEQ ID NO:2, may be isolated from native or from heterologous organisms or cells (including, but not limited to, bacteria, fungi, insect, plant, and mammalian cells) into which a protein-coding sequence has been introduced and expressed. Furthermore, the polypeptides may be part of recombinant fusion proteins.

Methods for polypeptide purification are well known in the art, including, without limitation, preparative discontiuous gel elctrophoresis, isoelectric focusing, HPLC, reversed-phase HPLC, gel filtration, ion exchange and partition chromatography, and countercurrent distribution. For some purposes, it is preferable to produce the polypeptide in a recombinant system in which the protein contains an additional sequence tag that facilitates purification, such as, but not limited to, a polyhistidine sequence. The polypeptide can then be purified from a crude lysate of the host cell by chromatography on an appropriate solid-phase matrix. Alternatively, antibodies produced against a protein or

against peptides derived therefrom can be used as purification reagents. Other purification methods are possible.

The present invention also encompasses derivatives and homologues of polypeptides. For some purposes, nucleic acid sequences encoding the peptides may be altered by substitutions, additions, or deletions that provide for functionally equivalent molecules, i.e., function-conservative variants. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of similar properties, such as, for example, positively charged amino acids (arginine, lysine, and histidine); negatively charged amino acids (aspartate and glutamate); polar neutral amino acids; and non-polar amino acids.

The isolated polypeptides may be modified by, for example, phosphorylation, sulfation, acylation, or other protein modifications. They may also be modified with a label capable of providing a detectable signal, either directly or indirectly, including, but not limited to, radioisotopes and fluorescent compounds.

The present invention encompasses antibodies that specifically recognize immunogenic components derived from C2/4GnT. Such antibodies can be used as reagents for detection and purification of C2/4GnT.

C2/4GnT specific antibodies according to the present invention include polyclonal and monoclonal antibodies. The antibodies may be elicited in an animal host by immunization with C2/4GnT components or may be formed by *in vitro* immunization of immune cells. The immunogenic components used to elicit the antibodies may be isolated from human cells or produced in recombinant systems. The antibodies may also be produced in recombinant systems programmed with appropriate antibody-encoding DNA. Alternatively, the antibodies may be constructed by biochemical reconstitution of purified heavy and light chains. The antibodies include hybrid antibodies (i.e., containing two sets of heavy chain/light chain combinations, each of which recognizes a different antigen), chimeric antibodies (i.e., in which either the heavy chains, light chains, or both, are fusion proteins), and univalent antibodies (i.e., comprised of a heavy chain/light chain complex bound to the constant region of a second heavy chain). Also included are Fab fragments, including Fab' and $F(ab)_2$ fragments of antibodies. Methods for the production of all of the above types of antibodies and derivatives are well known in the art. For example, techniques for producing and processing polyclonal antisera are disclosed in Mayer and

Walker, 1987, *Immunochemical Methods in Cell and Molecular Biology*, (Academic Press, London).

The antibodies of this invention can be purified by standard methods, including but not limited to preparative disc-gel elctrophoresis, isoelectric focusing, HPLC, reversed-phase HPLC, gel filtration, ion exchange and partition chromatography, and countercurrent distribution. Purification methods for antibodies are disclosed, e.g., in *The Art of Antibody Purification*, 1989, Amicon Division, W.R. Grace & Co. General protein purification methods are described in *Protein Purification: Principles and Practice*, R.K. Scopes, Ed., 1987, Springer-Verlag, New York, NY.

Anti C2/4GnT antibodies, whether unlabeled or labeled by standard methods, can be used as the basis for immunoassays. The particular label used will depend upon the type of immunoassay used. Examples of labels that can be used include, but are not limited to, radiolabels such as ^{32}P, ^{125}I, ^{3}H and ^{14}C; fluorescent labels such as fluorescein and its derivatives, rhodamine and its derivatives, dansyl and umbelliferone; chemiluminescers such as luciferia and 2,3-dihydrophthalazinediones; and enzymes such as horseradish peroxidase, alkaline phosphatase, lysozyme and glucose-6-phosphate dehydrogenase.

The antibodies can be tagged with such labels by known methods. For example, coupling agents such as aldehydes, carbodiimides, dimaleimide, imidates, succinimides, bisdiazotized benzadine and the like may be used to tag the antibodies with fluorescent, chemiluminescent or enzyme labels. The general methods involved are well known in the art and are described in, e.g., Chan (Ed.), 1987, *Immunoassay: A Practical Guide*, Academic Press, Inc., Orlando, FL.

Core 2 O-glycans are involved in cell-cell adhesion events through selectin binding, and the core 2 beta6GlcNAc-transferase activity is required for synthesis of the selectin ligands (11). The core 2 beta6GlcNAc-transferase activity therefore plays a major role in selectin mediated cell trafficking including cancer metastasis. Since at least two different core 2 synthases exist it is required to define which of these are involved in synthesis of O-glycans in different cell types and in disease. Development of inhibitors of individual or all core 2 synthase activities may be usefull in reducing or eliminating core 2 O-glycans in cells and tissues, and hence inhibiting the biological events these ligands are involved in. Inhibition of transcription and/or translation of core 2 beta6GlcNAc-transferase genes may have the same effect. Compounds with

such effects may be used as drugs with anti-inflammatory activity and/or for treatment of cancer growth and spreading.

The following examples are intended to further illustrate the invention without limiting its scope.

Example 1

A: Identification of cDNA homologous to C2/4GnT by analysis of EST database sequence information.

Database searches were performed with the coding sequence of the human C2GnT sequence () using the BLASTn and tBLASTn algorithms against the dbEST database at The National Center for Biotechnology Information, USA. The BLASTn algorithm was used to identify ESTs representing the query gene (identities of 95%), whereas tBLASTn was used to identify non-identical, but similar EST sequences. ESTs with 50-90% nucleotide sequence identity were regarded as different from the query sequence. Composites of all the sequence information for each set of ESTs were compiled and analysed for sequence similarity to human C2GnT.

B: Cloning and sequencing of C2/4GnT.

EST clone 178656 (5' EST GenBank accession number AA307800), derived from a putative homologue to C2GnT, was obtained from the American Type Culture Collection, USA. Sequencing of this clone revealed a partial open reading frame with significant sequence similarity to C2GnT. The coding region of human C2GnT and a bovine homologue was previously found to be organized in one exon (13) and unpublished observations). Since the 5' and 3' sequence available from the C2/4GnT EST was incomplete but likely to be located in a single exon, the missing 5' and 3' portions of the open reading frame was obtained by sequencing genomic P1 clones. P1 clones were obtained from a human foreskin genomic P1 library (DuPont Merck Pharmaceutical Co. Human Foreskin Fibroblast P1 Library) by screening with the primer pair TSHC27 (5'-GGAAGTTCATACAGTTCCCAC-3') (SEQ ID NO:3) and TSHC28 (5'-CCTCCCATTCAACATCTTGAG -3') (SEQ ID NO:4). Two genomic clones for C2/4GnT, DPMC-HFF#1-1026(E2) and DPMC-HFF#1-1091(F1) were obtained from Genome Systems Inc. DNA from P1 phage was prepared as recommended by Genome Systems Inc. The entire coding sequence of the C2/4GnT

gene was represented in both clones and sequenced in full using automated sequencing (ABI377, Perkin-Elmer). Confirmatory sequencing was performed on a cDNA clone obtained by PCR (30 cycles at 95°C for 15 sec; 55°C for 20 sec and 68° C for 2 min 30 sec) on total cDNA from the human COLO 205 cancer cell line with the sense primer TSHC54 (5'-GCAGAATTCATGGTTCAATGGAAGAGACTC-3') (SEQ ID NO:7) and the anti-sense primer TSHC45 (5'-AGCGAATTCAGCTCAAAGTTCAGTCCCATAG-3') (SEQ ID NO:5). The composite sequence contained an open reading frame of 1314 base pairs encoding a putative protein of 438 amino acids with type II domain structure predicted by the TMpred-algorithm at the Swiss Institute for Experimental Cancer Research (ISREC) (http://www.isrec.isb-sib.ch/software/TMPRED_form.html). The sequence of the 5'-end of C2/4GnT mRNA including the translational start site and 5'-UTR was obtained by 5' rapid amplification of cDNA ends (35 cycles at 94°C for 20 sec; 52°C for 15 sec and 72°C for 2 min) using total cDNA from the human COLO 205 cancer cell line with the anti-sense primer TSHC48 (5'-GTGGGAACTGTATGAACTTCC-3') (SEQ ID NO:6) (Fig. 2).

Example 2

A: Expression of C2/4GnT in Sf9 cells.

An expression construct designed to encode amino acid residues 31-438 of C2/4GnT was prepared by PCR using P1 DNA, and the primer pair TSHC55 (5'-CGAGAATTCAGGTTGAAGTGTGACTC -3') (SEQ ID NO:8) and TSHC45 (SEQ ID NO:5) (Fig. 2). The PCR product was cloned into the *Eco*RI site of pAcGP67A (PharMingen), and the insert was fully sequenced. Plasmids pAcGP67-C2/4GnT-sol and pAcGP67-C2GnT-sol were co-transfected with Baculo-Gold™ DNA (PharMingen) as described previously (14). Recombinant Baculo-virus were obtained after two successive amplifications in Sf9 cells grown in serum-containing medium, and titers of virus were estimated by titration in 24-well plates with monitoring of enzyme activities. Controls included the pAcGP67-GalNAc-T3-sol (15).

B: Analysis of C2/4GnT activity.

Standard assays were performed using culture supernatant from infected cells in 50 µl reaction mixtures containing 100 mM MES (pH 8.0), 10 mM EDTA, 10 mM 2-

Acetamido-2-deoxy-D-glucono-1,5-lacton, 180 μM UDP-[^{14}C]-GlcNAc (6,000 cpm/nmol) (Amersham Pharmacia Biotech), and the indicated concentrations of acceptor substrates (Sigma and Toronto Research Laboratories Ltd., see Table I for structures). Semi-purified C2/4GnT was assayed in 50 μl reaction mixtures containing 100 mM MES (pH 7), 5 mM EDTA, 90 μM UDP-[^{14}C]-GlcNAc (3,050 cpm/nmol) (Amersham Pharmacia Biotech), and the indicated concentrations of acceptor substrates. Reaction products were quantified by chromatography on Dowex AG1-X8.

<u>**Example 3**</u>

Restricted organ expression pattern of C2/4GnT

Total RNA was isolated from human colon and pancreatic adenocarcinoma cell lines AsPC-1, BxPC-3, Capan-1, Capan-2, COLO 357, HT-29, and PANC-1 essentially as described (17). Twentyfive μg of total RNA was subjected to electrophoresis on a 1% denaturing agarose gel and transferred to nitrocellulose as described previously (17). The cDNA-fragment of soluble C2/4GnT was used as a probe for hybridization. The probe was random primer-labeled using [α^{32}P]dCTP and an oligonucleotide labeling kit (Amersham Pharmacia Biotech). The membrane was probed overnight at 42°C as described previously (15), and washed twice for 30 min each at 42°C with 2 × SSC, 0.1% SDS and twice for 30 min each at 52°C with 0.1 × SSC, 0.1 % SDS. Human multiple tissue Northern blots, MTN I and MTN II (CLONTECH), were probed as described above and washed twice for 10 min each at room temperature with 2 × SSC, 0.1% SDS; twice for 10 min each at 55°C with 1 × SSC, 0.1 % SDS; and once for 10 min with 0.1 × SSC, 0.1 % SDS at 55°C.

<u>**Example 4**</u>

Genomic structure of the coding region of C2/4GnT

Human genomic clones were obtained from a human foreskin genomic P1 library (DuPont Merck Pharmaceutical Co. Human Foreskin Fibroblast P1 Library) by screening with the primer pair TSHC27 (5'-GGAAGTTCATACAGTTCCCAC-3') (SEQ ID NO:3) and TSHC28 (5'-CCTCCCATTCAACATCTTGAG -3') (SEQ ID NO:4). Two genomic clones for C2/4GnT, DPMC-HFF#1-1026(E2) and DPMC-HFF#1-1091(F1) were obtained from Genome Systems Inc. DNA from P1 phage was

prepared as recommended by Genome Systems Inc. The entire coding sequence of the C2/4GnT gene was represented in both clones and sequenced in full using automated sequencing (ABI377, Perkin-Elmer). Intron/exon boundaries were determined by comparison with the cDNA sequences optimising for the gt/ag rule (Breathnach and Chambon, 1981).

Example 5

Chromosomal localization of C2/4GnT: In situ hybridization to metaphase chromosomes

P1 DNA was labeled with biotin-14-dATP using the bio-NICK system (Life Technologies). The labeled DNA was precipitated with ethanol in the presence of herring sperm DNA. Precipitated DNA was dissolved and denatured at 80 C for 10 min followed by incubation for 30 min at 37 C and added to heat-denatured chromosome spreads where hybridization was carried out over night in a moist chamber at 37 C. After posthybridization washing (50% formamide, 2 x SSC at 42 C) and blocking with nonfat dry milk powder, the hybridized probe was detected with avidin-FITC (Vector Laboratories) followed by two amplification steps using rabbit-anti-FITC (Dako) and mouse-anti-rabbit FITC (Jackson Immunoresearch). Chromosome spreads were mounted in antifade solution with blue dye DAPI.

Example 6

Analysis of DNA polymorphism of C2/4GnT gene

Primer pairs as described in Figure 8 have been used for PCR amplification of individual sequences of the coding exon III. Each PCR product was subcloned and the sequence of 10 clones containing the appropriate insert was determined assuring that both alleles of each individual are characterized.

From the foregoing it will be evident that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

REFERENCES

1. Clausen, H. and Bennett, E.P. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. Glycobiology, *6*: 635-646, 1996.

2. Piller, F., Piller, V., Fox, R.I., and Fukuda, M. Human T-lymphocyte activation is associated with changes in O-glycan biosynthesis. J.Biol.Chem., *263*: 15146-15150, 1988.

3. Yang, J.M., Byrd, J.C., Siddiki, B.B., Chung, Y.S., Okuno, M., Sowa, M., Kim, Y.S., Matta, K.L., and Brockhausen, I. Alterations of O-glycan biosynthesis in human colon cancer tissues. Glycobiology, *4*: 873-884, 1994.

4. Yousefi, S., Higgins, E., Daoling, Z., Pollex-Kruger, A., Hindsgaul, O., and Dennis, J.W. Increased UDP-GlcNAc:Gal beta 1-3GalNAc-R (GlcNAc to GalNAc) beta-1, 6- N-acetylglucosaminyltransferase activity in metastatic murine tumor cell lines. Control of polylactosamine synthesis. J.Biol.Chem., *266*: 1772-1782, 1991.

5. Fukuda, M. Possible roles of tumor-associated carbohydrate antigens. Cancer Res., *56*: 2237-2244, 1996.

6. Brockhausen, I., Yang, J.M., Burchell, J., Whitehouse, C., and Taylor-Papadimitriou, J. Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. Eur.J.Biochem., *233* : 607-617, 1995.

7. Brockhausen, I., Kuhns, W., Schachter, H., Matta, K.L., Sutherland, D.R., and Baker, M.A. Biosynthesis of O-glycans in leukocytes from normal donors and from patients with leukemia: increase in O-glycan core 2 UDP-GlcNAc:Gal beta 3 GalNAc alpha-R (GlcNAc to GalNAc) beta(1-6)-N- acetylglucosaminyltransferase in leukemic cells. Cancer Res., *51*: 1257-1263, 1991.

8. Higgins, E.A., Siminovitch, K.A., Zhuang, D.L., Brockhausen, I., and Dennis, J.W. Aberrant O-linked oligosaccharide biosynthesis in lymphocytes and platelets from patients with the Wiskott-Aldrich syndrome. J.Biol.Chem., *266*: 6280-6290, 1991.

9. Saitoh, O., Piller, F., Fox, R.I., and Fukuda, M. T-lymphocytic leukemia expresses complex, branched O-linked oligosaccharides on a major sialoglycoprotein, leukosialin. Blood, *77*: 1491-1499, 1991.

10. Springer, G.F. T and Tn, general carcinoma autoantigens. Science, *224*: 1198-1206, 1984.

11. Kumar, R., Camphausen, R.T., Sullivan, F.X., and Cumming, D.A. Core2 beta-1,6-N-acetylglucosaminyltransferase enzyme activity is critical for P-selectin glycoprotein ligand-1 binding to P-selectin. Blood, *88*: 3872-3879, 1996.

12. Bierhuizen, M.F. and Fukuda, M. Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen. Proc.Natl.Acad.Sci.U.S.A., *89*: 9326-9330, 1992.

13. Bierhuizen, M.F., Maemura, K., Kudo, S., and Fukuda, M. Genomic organization of core 2 and I branching beta-1,6-N- acetylglucosaminyltransferases. Implication for evolution of the beta- 1,6-N-acetylglucosaminyltransferase gene family. Glycobiology, *5*: 417-425, 1995.

14. Almeida, R., Amado, M., David, L., Levery, S.B., Holmes, E.H., Merkx, G., van Kessel, A.G., Rygaard, E., Hassan, H., Bennett, E., and Clausen, H. A family of human beta4-galactosyltransferases. Cloning and expression of two novel UDP-galactose:beta-n-acetylglucosamine beta1, 4- galactosyltransferases, beta4Gal-T2 and beta4Gal-T3. J.Biol.Chem., *272* : 31979-31991, 1997.

15. Bennett, E.P., Hassan, H., and Clausen, H. cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D- galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3. J.Biol.Chem., *271*: 17006-17012, 1996.

16. Wandall, H.H., Hassan, H., Mirgorodskaya, E., Kristensen, A.K., Roepstorff, P., Bennett, E.P., Nielsen, P.A., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. Substrate specificities of three members of the human UDP-N-acetyl- alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. J.Biol.Chem., *272*: 23503-23514, 1997.

17. Sutherlin, M.E., Nishimori, I., Caffrey, T., Bennett, E.P., Hassan, H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M.A. Expression of three UDP-N-acetyl-alpha-D-galactosamine:polypeptide GalNAc N-acetylgalactosaminyltransferases in adenocarcinoma cell lines. Cancer Res., *57*: 4744-4748, 1997.